Faculty of Electrical Engineering
Universiti Teknologi Malaysia
VOL. 10, NO. 1, 2008, 60-67
ELEKTRIKA
http://fke.utm.my/elektrika

An Acceleration Simulation Method for Power law Priority Traffic

Sharifah H. S. Ariffin[1*] and **John A. Schormans**[2]

[1]Faculty of Electrical Engineering, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor, Malaysia.
[2]Department of Electronic Engineering, Queen Mary, University of London, London, UK.

[*]Corresponding author: sharifah@fke.utm.my (Sharifah S. H. Ariffin), Tel: 607-5535348, Fax: 607-5566262

Abstract: A method for accelerated simulation for simulated self-similar processes is proposed. This technique simplifies the simulation model and improves the efficiency by using excess packets instead of packet-by-packet source traffic for a FIFO and non-FIFO buffer scheduler. In this research is focusing on developing an equivalent model of the conventional packet buffer that can produce an output analysis (which in this case will be the steady state probability) much faster. This acceleration simulation method is a further development of the Traffic Aggregation technique, which had previously been applied to FIFO buffers only and applies the Generalized Ballot Theorem to calculate the waiting time for the low priority traffic (combined with prior work on traffic aggregation). This hybrid method is shown to provide a significant reduction in the process time, while maintaining queuing behavior in the buffer that is highly accurate when compared to results from a conventional simulation.

Keywords: Accelerate simulation, Generalized ballot theorem, Power law, Self-similar traffic

1. INTRODUCTION

The enormous growth of network users and the rich class of applications in modern IP networks have made network performance analysis more complex. With data, voice and videos streaming over the packet switched network, network users expect certain Quality of Service (QoS). Previously observation that telephone networks traffic followed the Poisson distribution was widely used. This distribution implies that call arrivals to the networks are independent and the call inter-arrival times are exponentially distributed. Empirical studies of measured traffic in modern packet networks have led to the wide recognition of the self-similar nature of the measured aggregate network traffic and the underlying heavy-tailed distributions [1-6]. The changing characteristics of the modern traffic networks show strong evidence of persistent burstiness regardless of timescales unlike the low variation as in voice traffic.

Simulation is used most for network performance analysis, to estimate planning costs and congestions. Analytical methods serve the same purpose, but problems arise as networks become larger, more complex with variety of QoS priority sources. For these reasons, in recent years, many researchers have focused on modeling realistic telecommunication network with complex traffic models. Self-similar and long-range dependence, which are the important features of aggregated traffic, reflect the real statistical behavior of modern network traffic. Conventional simulation methods, i.e. at the packet level, can keep track of individual packets as they travel across the networks but they have drawbacks such as the large computational requirements (both in processing and storage) for large-scale simulation. As a consequence, with conventional simulation methods simulating a few hours of simulation time might take days of real time.

This paper propose an accelerated simulation method using aggregated technique that supports different types of traffic flows buffered by a non-FIFO scheduler and based on the previously reported Traffic Aggregation (TA) method [12]. Using this approach the queuing behavior of self-similar arrival process, specifically those with Pareto distributed activity periods. We apply this approach to homogeneous traffic using FIFO scheduler and heterogeneous traffic sources with non-FIFO scheduler. Both scenarios are verified with the packet-by-packet simulation method. The paper focus on the key issues of simulation acceleration in the modern packet switched network: rare events occurrence i.e. buffer overflow probabilities and the real time taken for the simulation.

This paper is organized as follows: Section 2 discusses related work in accelerated simulation. The paper presents a brief introduction to ON/OFF sources with heavy-tailed distribution, namely Pareto distribution in section 3 as well as the 4 proposes the accelerated simulation method with FIFO scheduler and non-FIFO scheduler. Section 4 shows the results and discussion. The conclusions are in section 5.

2. RELATED WORK

Accelerated simulation methods can be categorized into the following types a) computational power, b) simulation technology and c) simulation model [15]. Computational power i.e. parallel simulation speeds up simulations using sub-models 3 with an existing sequential simulation tool using the full expressive power of the tool [19-21]. Simulations can be further accelerated by the second category where new enhanced algorithms such as variance reduction are applied. The simulation technology uses statistical methods to obtain more accurate performance analysis within a reasonable time. Examples that had received much attention in the literature are the Importance Sampling [7-9, 22, 23] and RESTART methods [24]. The third category, which is the simulation model, simplifies the simulation model and improves its efficiency. A hybrid method as in TA is an example in this category that combine analytical models with a simulation model [11, 25, 27]. In the 70's hybrid simulations were developed as less expensive version of discrete event simulations model of computer systems [26] but nowadays hybrid method is used in all types of network simulation to increase efficiency and reduce simulation time [16, 10-14]. The limitation of these acceleration models is that most situation examples are done in homogeneous traffic environment, so in this paper we proposed an approach to model the packet switched network with aggregate packet-train traffic supporting heterogeneous traffic offered by the real modern network provider. This method uses hybrid method that involves partitioning the buffer queue into sub queues with prior knowledge of the mean busy period distribution in the high priority sub queue.

3. ACCELERATION TECHNIQUE

3.1 Traffic Aggregation (TA)

Studies of queuing theory in [28, 29] have discovered that a single 2 state process can replace N x 2 state processes accurately. The resulting 2 state processes are either in the ON state, or in the OFF state. Using this traffic aggregation any N multiplexed sources can be replaced by a single equivalent ON/OFF source with equivalent queuing behaviour in the buffer. This concept has been successfully applied to Markovian ON/OFF sources [25, 30] and traffic sources with power law sojourn times [10, 11, 13, 14, 31]. In this paper aggregation involving power law traffic will be what we refer to as Traffic Aggregation (TA.).

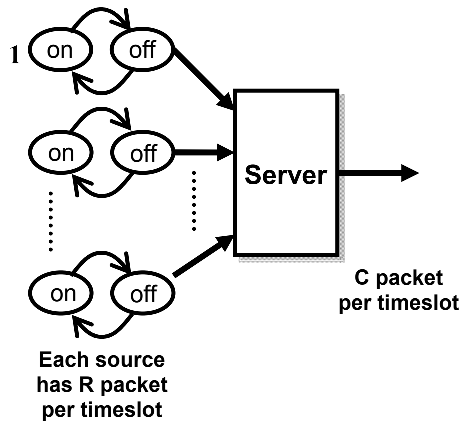
Consider a conventional (conventional is define as simulator without acceleration) traffic model, comprising N ON-OFF sources with Pareto distributed sojourn times (see Figure 1a). These individual traffic sources transmit packets at a rate of R packet/timeslot during the ON periods and R=0 during the OFF periods. The mean ON and OFF periods are denoted by *Don* and *Doff* respectively. The mean arrival rate for an individual source is given by:

$$\lambda_{Conv} = R \cdot \alpha \tag{1}$$

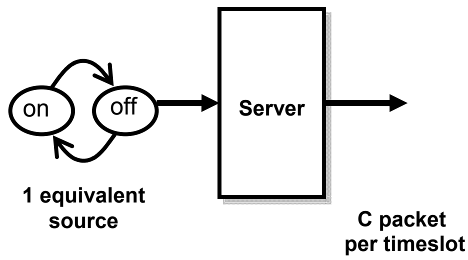
$$\alpha = \frac{Don}{Don + Doff} \tag{2}$$

where α is the probability that the individual source is active. In this work queuing system use a deterministic service rate, which transmits packet at C packets/timeslot (C is usually assume equal to 1). The randomly switching of states (ON to OFF and vice versa) from this individual source results in a traffic pattern that is similar to the illustration of Figure 1c. Random switch here means the decision to be in ON or in OFF state is determined by a random simulation generator (RNG). In this case the sojourn ON time is not random (i.e. exponential) but Pareto distributed. The First-In-First-Out (FIFO) queuing mechanism is used and the utilization of the system, ρ, is given by:

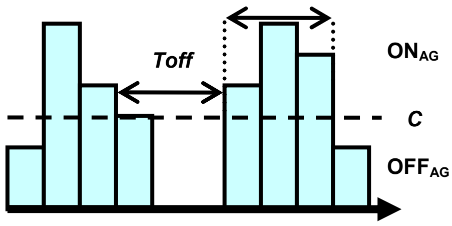
$$\rho = \alpha \cdot N \cdot R \cdot \frac{1}{C} \tag{3}$$



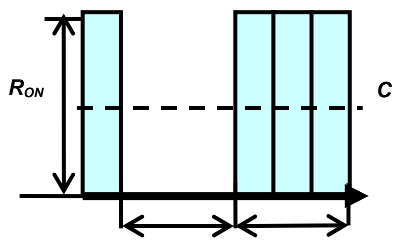
(a) Conventional: N traffic sources



(b) TA: 1 equivalent



(c) Conventional: resultant traffic



(d) TA: aggregated traffic

Figure 1. The number of events had been reduce in TA

TA introduced an acceleration technique that reduces the multiplexed scenario to only one single equivalent source (see Figure 1b.). The aggregated 2 state process of the equivalent 2 state model is either in the ON state or the OFF state; these are defined by reference to the resultant traffic pattern from the N original sources (see Figure 1c). The ON and OFF states of the aggregated processes are denoted ON_{AG} and OFF_{AG} respectively, and these (ON and OFF states) are derived with reference to the service rate, *C*. An ON_{AG} period occurs when the packet arrival rate is greater than the service rate, i.e. *NR*>*C*, and the mean duration of this aggregated state is denoted by *Ton*. An OFF_{AG} period occurs when there is a continuous number of timeslots in which the packet arrival rate is less than or equal to the service rate, and the mean of this duration is denoted as *Toff*. Figure 1d shows that the rate of packet arrivals during the OFF period is always zero (refer to for detail), and therefore the accumulated. The packet arrival rate that is equal to the service rate, *NR*=*C*, is considered as OFF because no effect on the buffer size, and no further accumulation of packets.

3.2 The Concept of Enhanced Traffic Aggregation for FIFO Scheduler (E_TA-FIFO)

In E_TA, the excess rate (ER) arrivals represent the excess of arrivals during the ON period of the aggregated source (i.e. those arrivals that force a queue to build up in the buffer). [32] gives the detail explanation of excess packets. In a packet-by packet model (i.e. conventional), a simulation event occurs when there is an arrival of a packet, but in E_TA a simulation event occurs only when there is a change of state, i.e. from the ON state to the OFF state or vice versa. Moving from state OFF to ON in E_TA will add an 'ER batch' to the queue. This forms an acceleration technique that is more advanced, and requires fewer simulated events than TA. For example, in TA a particular ON period might send M packets and the total number of events to complete one cycle of ON/OFF periods is M + 1 (including the end of the OFF period) events. Using E_TA, the total number of events in 1 ON/OFF cycle is 2 (the end of the ON period and the end of the OFF period), this is regardless of the number of packets in the ON period. The definition of events in E_TA and TA is illustrated in Figure 2.

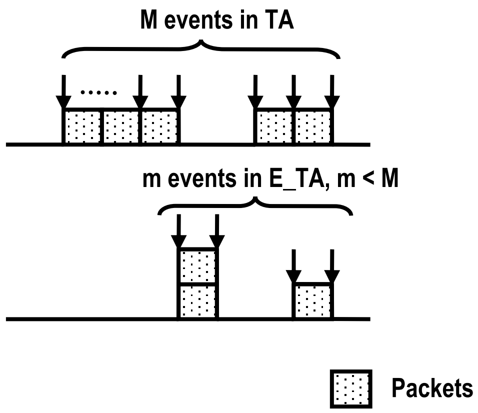


Figure 2. Simulation events in packet-by-packet model (i.e. TA) and E_TA model.

The mean size of an ER batch is given by:

$$\overline{ER} = Ton(Ron - C) \tag{4}$$

where Ton is the mean ON period of the aggregated traffic, Ron is the mean arrival rate in the ON period of the aggregated traffic and C is a service rate (packet/ timeslot).

The mean OFF period in E_TA, $Toff_2$, which is also the inter-arrival time of the ER batches, is found from the following equation:

$$Toff_2 = \frac{\overline{ER}}{\rho \cdot C} \tag{5}$$

where ρ is the utilization of the system. This is because ON periods in E_TA last for 1 timeslot.

As already described previously an E_TA simulation model is divided into slots where each is of duration equal to the fixed service time of a packet. In a conventional packet-by-packet simulator, if the service rate is C, the service time of each individual packet will be 1/C. However in E_TA, Pareto distributed batches of ER arrivals will arrive in any slot with probability *q* and the FIFO buffer holds ER batches instead of individual packets. Hence the mean service time of each ER batch in the FIFO buffer, $\overline{ST_B}$, is given by:

$$\overline{ST_B} = \frac{\overline{ER}}{C} \tag{6}$$

3.3 The Concept of Enhanced Traffic Aggregation for non FIFO Scheduler (E_TA-PQ)

The E_TA algorithm for a priority scheduler supports two priority levels; the priority type is non-pre-emptive, because all real packet-scheduling systems are non-pre-emptive. The high priority traffic is set to represent VoIP, and, as an aggregate of Constant Bit Rate voice traffic streams (for EF PHB), is well modeled by a Poisson process [33], we set the high priority traffic model to be Poisson. In contrast, the low priority traffic is for AF and BE PHB groups, which includes: FTP, TELNET and SMTP (email) traffic. Studies [5] have shown that over larger time scales most of the AF and BE traffic appears to be bursty, and dominated by self-similar characteristics.

The scheduler in PQ works as follows: when all the packets in the busy period of the high priority sub queue are served, the server will switch to low priority sub queue. Low priority packets may arrive after the busy period of high priority ends, in which case these packets are served right away (when the high priority sub queue is idle). However, if low priority packets arrive during a busy period of high priority, then these packets will have to wait until the busy period ends.

Studies of queuing analysis, e.g. in [35, 36, 29, 28] used the powerful GBT [34] to analyze the steady state waiting time probabilities in PQ through busy period analysis. In the PQ, the waiting time of the low priority packets depends very much on the high priority traffic, (see Figure 3). The waiting time for any packet is defined as the total time from entry to the system to entry to the server (being a non-pre-emptive scheduler, a packet will then complete transmission after its service time with probability 1).

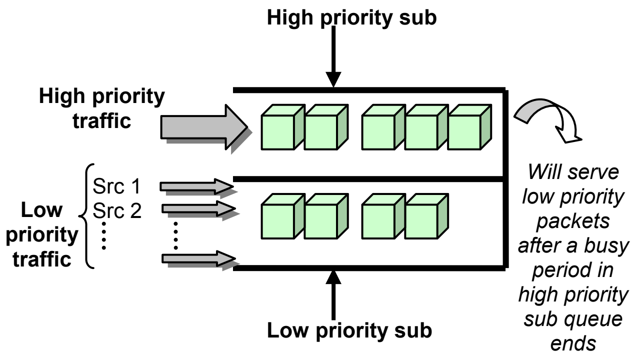


Figure 3. The priority buffer of two priority levels

The mean service time of each ER batch in a FIFO scheduler is given by equation (4). For the E_TA-PQ model, the service time of the low priority traffic is equivalent to the total time each ER batch spends in the buffer. To find the waiting time probabilities of the low priority traffic requires the use of convolution. Based on the waiting time probability for two priority levels in [29], a packet or in this case an ER batch, say P , arriving in timeslot i, will have to wait behind a number of low priority ER batches and high priority packets. This total wait has 3 essential

1. the total number of packets of equal or higher priority that were already present in the buffer at the end of timeslot *i*-1
2. all the packets of higher priority that ***P*** that arrive in timeslot *i*
3. higher priority packets that arrive subsequently, but before ***P*** enters service.

Define component 1 — the unfinished work- as *u (k),*
component 2 — the wait caused by high priority arrival in timeslot *i*,
component 3 — the extra wait cause by the subsequent arrivals of the high priority traffic.

3.3 The Hybrid Approach for E_TA-PQ

The waiting time probability of the low priority traffic consist of

1. the service time of ER batch itself.
2. the extra wait cause by subsequent arrivals of the high priority traffic.

In this research, * denotes the convolution operator and all packet sizes are fixed.

Definitions:
High Priority will be denoted 1 and Low Priority 2
$a_1(k)$ = Pr (k packets of high priority arrive in any timeslot)
$a_2(k)$ = Pr (k packets of low priority arrive in any timeslot)
$a_c(1,k)$ = Pr (k units of work of high or low priority arrive in any 1 timeslot)
ρ_1 = load of high priority traffic
ρ_2 = load of low priority traffic
$u(k)$ = Pr (an arriving batch of low priority packets sees k packets of unfinished work of high or low priority already in the queue)
$V_2(k)$ = Pr (a low priority batch waits k timeslots due to the unfinished work of other high and low priority arrivals already ahead of it when it joins the queue)
$W_2(k)$ = Pr (a low priority batch has a total wait of k timeslots)

Due to our model of VoIP as priority 1, the high priority traffic is Poisson, so $a_1(k)$ is simply:

$$a_1(k) = \frac{\rho_1^{\,k}}{k!} \cdot e^{-\rho_1} \tag{7}$$

The probability of low priority batch arrivals is a discrete distribution whereas the Pareto is a continuous distribution. Using the discrete-time queuing model for long-range dependent (LRD) traffic in [32], the probability that a Pareto batch is size of *k* can be found using.

$$gp(k) = \left(\frac{1}{k-0.5}\right)^{\alpha} - \left(\frac{1}{k+0.5}\right)^{\alpha} \tag{8}$$

and $\alpha = \frac{B}{B-1}$

where *B* is the mean size of the ER batch arrivals in the E_TA model. This gives the probability that there is an ER batch (of size *k*) arrivals in any timeslot as:

$$a_2(k) = \begin{cases} 1-q & k=0 \\ q \cdot gp(k) & k>0 \end{cases} \tag{9}$$

$$q = \frac{\rho_2}{B}$$

where q is the probability of there being a batch in any random timeslot.

Because the distribution of the high and low priority traffic are different the unfinished work, $u(k)$, is actually found from the convolution of the distribution of the high and low priority traffic which gives:

$$a_c(1,k) = \begin{cases} a_1(0) * a_2(0) & k=0 \\ a_1(k) * a_2(k) & k>0 \end{cases} \tag{10}$$

$$s_2(k) = \begin{cases} s_2(0) \cdot \dfrac{1-a_c(1,k)}{a_c(1,0)} & k=1 \\ \dfrac{s_2(k-1) - s_2(0)\cdot a_c(1,k-1) - \sum_{i=0}^{k-1} s_2(i)\cdot a_c(1,k-i)}{a_c(1,0)} & k>1 \end{cases} \tag{11}$$

where $s_2(0) = 1 - E[a]$, $E[a]$ is the mean number of arrivals per timeslot of both high and low priority arrivals.

$$E[a] = \rho_1 + \rho_2 \tag{12}$$

Hence,

$$u(k)=\begin{cases} s_2(0)+s_2(1) & k=0 \\ s_2(k+1) & k>0 \end{cases} \tag{13}$$

The virtual waiting time for the low priority batch can be found using

$$V_2(k)=a_1(k)*u(k) \tag{14}$$

as all two distributions are independent.

The GBT is applied to find the extended waiting time caused by subsequent arrivals of high priority packets. Stated in the form that is applied, the GBT is:

$$pr\{y=k|x=i\}=\frac{i}{k}\cdot(pr(k-i)) \tag{15}$$

where y is a number of packets in the busy period of high priority packets, x is a number of packets initially in the system and $pr(k-i)$ is Probability that k high priority packets arrive in (k-i) timeslots.

The waiting time probability for the low priority batch is found using:

$$W_2(k)=\begin{cases} V_2(0) & k=0 \\ \sum_{i=1}^{k}\left[V_2(i)\cdot\frac{i}{k}\cdot pr(k-i)\right] & k>0 \end{cases} \tag{16}$$

where $W_2(k)=$ Pr (a low priority batch must wait k timeslots before it enters service).

The waiting time probability of the low priority traffic, $W_2(k)$, is calculated before any simulation is executed. All values of $W_2(k)$ are written in the PDF editor table of OPNET. In the E_TA-PQ simulation, when an ER batch arrives, the simulation will randomly generate a value using implementation of $W_2(k)$ in OPNET.

4. RESULTS AND DISCUSSIONS

The concept of the ETA-FIFO method is first validated for a single node network. In the simulation active probes are use, which are generated at the sending end observation of the delay distributions at the receiving end. The steady state probability of packet delay is defined as:

Pr (*t*) = Pr (the delay of a packet probe is *t* timeslot)

Individual parameters of Don and N are varied, and the details of the simulations are listed in Table 1 and 2. The traffic values in these tables are chosen mainly because these values can be generated accurately (due to restricted range of random numbers generated by the Pareto distribution), and not specifically to represent any particular 'real' traffic or application type. E_TA-FIFO is tested as the number of the aggregated traffic sources is varied. Figure 4 shows that the packet delay probabilities of E_TA -FIFO are accurate compare to conventional and TA. Moreover better coverage of the probabilities was achieved in the same length of time.

Table 1. The parameters values for E_TA, TA and conventional model used in Figure 4

Conventional				
Doff	Don	N	R	C
10	3.1	10	1	5
10	3.9	10	1	5
10	5.0	10	1	5
TA				
Toff	Ton	Ron	C_{TA}	
2.3258	1.8059	6.2258	5	
2.2292	2.3086	6.2871	5	
2.2022	2.9998	6.3569	5	
E-TA				
$\overline{ER}$	$Toff_2$	C_{ETA}		
2.2137	0.9354	5		
2.9714	1.2556	5		
4.0704	1.7200	5		

Table 2. The individual and equivalent traffic sources value used in Figure 5

CONVENTIONAL				
Doff	Don	N	R	C
10	3.9	16	1	8
10	3.9	18	1	8
10	3.9	20	1	8
TA				
Toff	Ton	Ron	C_{TA}	
2.0658	1.9218	9.3149	8	
2.1551	2.4844	9.94314	8	
2.3774	3.3764	9.5627	8	
E_TA				
$\overline{ER}$	$Toff_2$	C_{ETA}		
2.5369	0.5628	8		
3.5562	0.7042	8		
5.2763	0.9403	8		

For E-TA PQ the packet delay of the low priority traffic will be of interest. Active measurement is used to collect the packet delay probability of the low priority sub queue. The probes are generated at the sending end and the delay probability is observed at the receiving end of the system. The arrival rate of the probe is set to be 0.001-probe/ timeslot. This low arrival rate is essential in order for the system to work without any significant interruption from the active probes.

The parameter values of the conventional and E_TA-PQ low priority traffic are shown in Table 3 and 4. The high priority is Poisson and the arrival rate, λ_1, is equal to 0.2 packet/timeslot. The graph compares the conventional N source with E_TA-PQ. General we can see that simulation-using E_TA-PQ with fewer events can reach lower probabilities in the same length of time. Figure 6 shows that the E_TA-PQ model has reproduced the packet delay accurately for different mean ON period of the individual low priority traffic. These results are for aggregated of 4 low priority traffic sources. Figure 7 shows the distribution of the packet delay for aggregated of 6 low priority traffic sources. Here, the mean ON period is varied and the queuing behavior is observed. Again it shows accuracy and more coverage in the time delay.

The queuing behavior is observed at the end of the ON period of E_TA-PQ, due to the modeling structure of ER batch arrivals this is because, the queuing behavior observed at the end of the ON period for E_TA is best compared to the one collected from TA at an equivalent point. Therefore, for the experiment in this section a packet-by-packet, PQ model is developed which has aggregated low priority traffic and this is referred to as TA1. As same as in the previous section, in this section we also focus exclusively on the low priority traffic of Figure 6 and 7. The steady state probability is defined as

Prlow(k) = Pr (k packets in the low priority sub queue end of the ON period)

The high priority traffic is set to have mean arrival rate of 0.2 packet/ timeslot. Two sets of results are presented. One is E_TA-PQ with 4 low priority traffic sources and the other one is with 6 sources of low priority traffic. Figure 6 shows the reproduction of the queuing behavior in the low priority traffic is very accurate compared to the one from TA-PQ model. Figure 7 also shows that E_TA-PQ can reproduce accurate queuing behavior.

Then the processing time taken for N events for TA-PQ and E_TA_PQ is compared in Figure 8. All OPNET simulations are terminated using end-simulation interrupts which are delivered to all processors and queues that have the "endsim intrpt" attribute enabled. In Figure 8, it clearly shown the significant reduction of processing time using the E_TA-PQ model compared to TA-PQ model.

Table 3. The parameter values of the low priority traffic used in Figure 6

Doff	Don	N	R	C	$\overline{ER}$	$Toff_2$	Ron	C_{ETA}
10	3.1	4	1	2	2.6904	2.8421	3.0719	2
10	3.5	4	1	2	2.8852	2.7821	3.0805	2
10	3.9	4	1	2	3.0821	2.7462	3.0888	2

Table 4. The individual and equivalent traffic sources value used in Figure 7.

Don	Don	N	R	C	$\overline{ER}$	$Toff_2$	Ron	C_{ETA}
10	3.1	6	3	3	2.5188	1.7740	4.1210	3
10	3.3	6	3	3	2.6455	1.7770	4.1286	3

Figure 4 shows the validation of the individual sub queues (high priority sub queue and low priority sub queue) in E_TA with non FIFO scheduler with mean ON duration, T_{on} = 8 unit time, OFF duration time, T_{off} = 3 unit time, ON arrival rate, R_{on} = 2 unit time and service rate of 1 unit time. Both validations are to ensure the operation of each sub queue independently of the other. Even with Markovian sources the sub queue shows accurate validation against the theoretical queue distribution, which is the steady state probability at the packet arrival instant, is evaluated defined in [28]. The example with mean ON duration, T_{on} = 10 unit time, OFF duration time, T_{off} = 3 unit time, ON arrival rate, R_{on} = 2 unit time and service rate of 1 unit time is shown in Figure 5.

In Figures 6 and 7, comparison between E_TA with non-FIFO scheduler and the original of packet-by-packet version gives accurate queue distribution for the low priority sub queue. In the examples with total loads of 0.5 and 0.46 shows that E_TA with non-FIFO scheduler can accurately reproduce the queuing behavior of the original packet-by-packet system. Moreover the wall clock simulation time was 83% accelerated using E_TA with non-FIFO scheduler model compared to the original packet-by-packet version. This is a further reduction of time compare to E_TA with FIFO scheduler.

5. CONCLUSIONS

In this paper we developed a method (E_TA) to accelerate simulation that can be used to simulate homogeneous and heterogeneous self-similar traffic in packet switched networks. This method is for both FIFO scheduler and a non-FIFO scheduler and a speed up in real time was achieve for both methods, but more acceleration was obtained from E_TA with non FIFO scheduler. The simulations experiments show that acceleration of simulation for self-similar traffic is possible for homogeneous and heterogeneous environment.

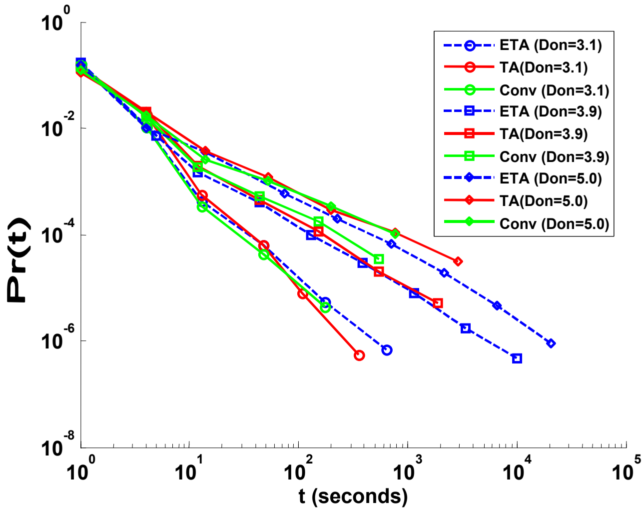


Figure 4. Probability of the packet delay for different values of Don

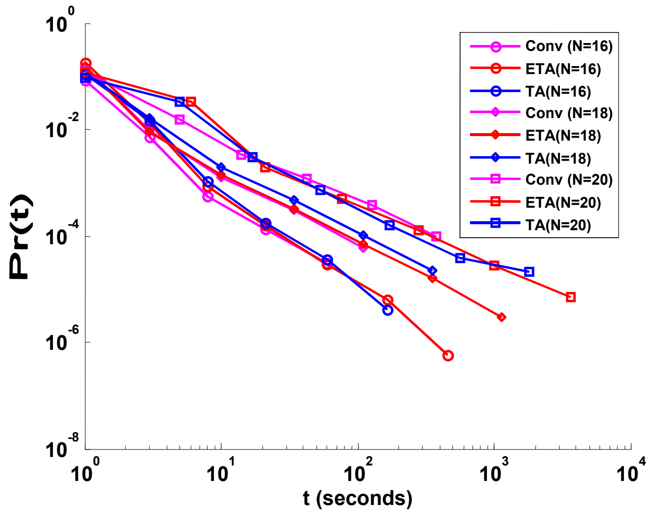


Figure 5. The probability of the packet delay for Different number of N

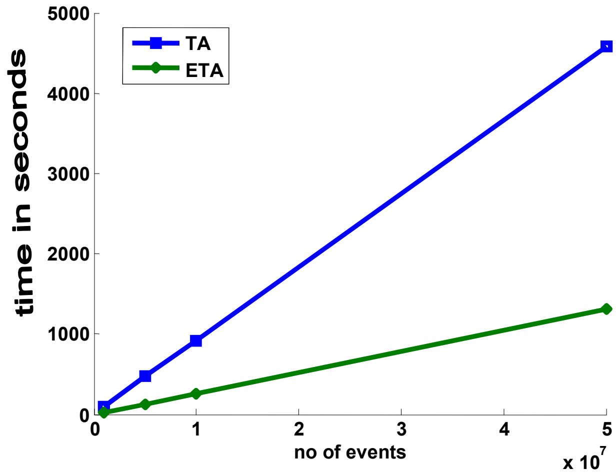


Figure 8. Time reduction in E_TA-PQ

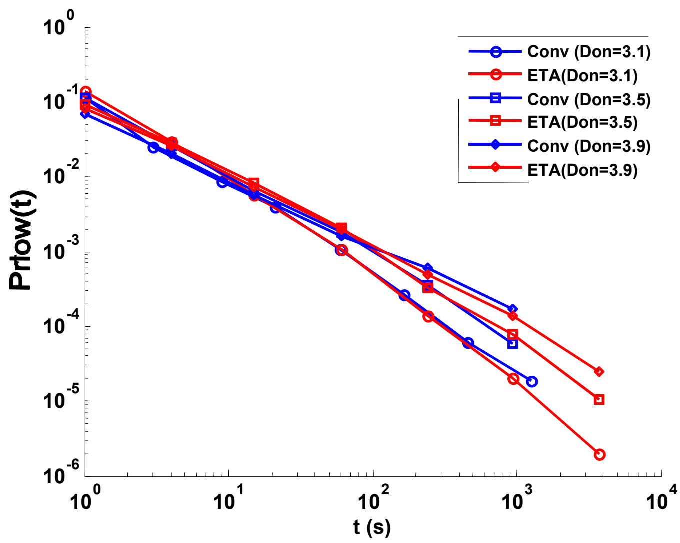


Figure 6. The PMF of low priority traffic packet delay (N=4)

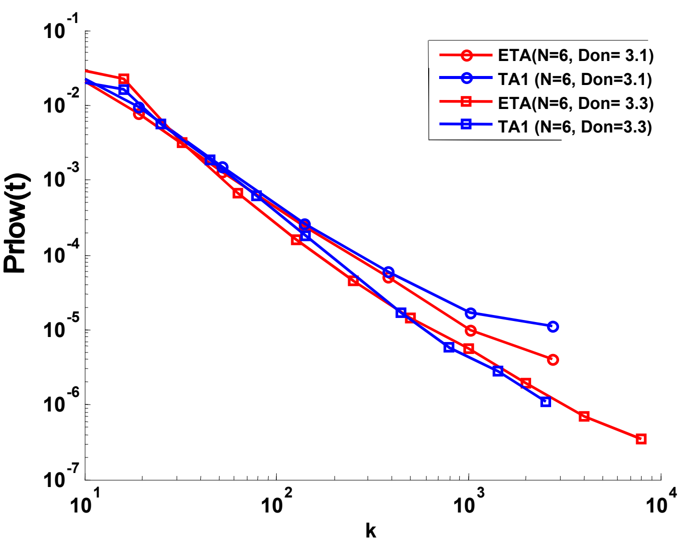


Figure 7. The queuing behavior with 6 the low priority traffic sources

REFERENCES

[1] J. Beran et al., “Long-Range Dependence in Variable Bit rate Video Traffic”, IEEE Transactions on Communication, 1995.

[2] M. Garret W. Willinger, “Analysis Modelling and Generation of Self Similar VBR Video Traffic”. *Proceedings of ACM SIGCOMM*, 1994.

[3] W. Leland, M. Taqqu, W. Willinger and D. Wilson, “On the Self-Similar Nature of Ethernet Traffic (Extended version)”, *IEEE/ ACM Transaction on Networking*, February 1994.

[4] W. Willingger, M. S. Taqqu, R. Sherman and D.V. Wilson, “Self-Similarity through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level”, *IEEE/ACM Transactions on Networking*, Vol.5, No. 1, February 1997.

[5] PAXSON V. and FLOYD S.: ‘Wide Area Traffic: The Failure of Poisson Modelling’, *IEEE/ACM Transactions on Networking*, 1995, 3, (3), pp.226-244.

[6] M.E. Crovella and M. Taqqu, “Estimating The Heavy Tail Index From Scaling Properties”, *Methodology and Computing in Applied Probability*, Vol. 1, No.

[7] J. R. Gallardo, D. Makrakis and L. Orozco-Barbosa, “Fast Simulation of Broadband Telecommunications Networks Carrying Long-Range Dependent Bursty Traffic”, Proc. of 1990 *Winter Simulation Conference*.

[8] J. K. Townsend, Z. Harastzi, J. A. Freebersyser and M. Devetsikiotis, “Simulation of Rare Events in Communication Networks”, *IEEE Comm*., Vol. 36, No. 8, August 1998.

[9] C. Huang, M. Devetsikiotis, I. Lambadaris and A. R. Kaye, “Fast Simulation for Self-Similar Traffic in ATM Networks”, *ICC95*, Vol. 1.

[10] A. Ma, J. Schormans and L. Cuthbert, “Aggregation Technique for Networks with Power Law Traffic and Application to accelerated Simulation”, *IEE Proc. Comm*., Vol. 150, No. 3, June 2003.

[11] A.H.I. Ma and J.A. Schormans, “Hybrid Technique for Analysis of Multiplexed Power-Law Traffic in

Broadband Networks”, *IEE Electronic Letters*, Vol.38, No. 6, March 2002.

[12] A. H. I. MA and J. A. Schormans , "A Fast Simulation Method For Modelling IP Networks", *17th UK Teletraffic Symposium on "Networks: A service centred approach*", Dublin, Ireland, May 2001

[13] A. H. I. MA and J. A. Schormans.: “Efficient Simulation Modelling for Long Range Dependent Traffic”, *Fourth Workshop on Rare Event Simulation and related Combinatorial Optimization Problems (RESIM / COP'02),* April 11-12, 2002, Madrid, Spain.

[14] A. H. I. MA and J. A. Schormans: “Accelerated Simulation Modelling Of Power-Law Traffic Via Aggregation*”, ITC2002* Beijing, June 2002.

[15] B. Liu, D.R. Figueiredo, Y Guo, J. Kurose and D. Townley, “A study of Networks Simulation Efficiency: Fluid Simulation Vs Packet Level Simulation”, *IEEE INFOCOM 2001*, Vol. 3.

[16] S.H.S. Ariffin and J.A. Schormans, “Accelerated Simulation Method for Power-law Traffic and Non-FIFO scheduling”, proceedings of *London Comm. Symposium 2003*, LCS2003.

[17] B. Mandelbrot, “Self-Similar Error Clusters in Communication Systems and the concept of conditional stationary”, *IEEE Transactions on Communication Technology* COM-13 (1965).

[18] M. Taqqu and J. Levy, “Using Renewal Processes to generate Long Range Dependence and High Variability”, *Dependence in Probability and Statistics* (Boston, MA, 1986)

[19] D. Nicol and P. Heidelberger, “Parallel execution for serial simulators”, *ACM Trans. On Modelling and Computer Simulation*, Vol. 6, No. 3, July 1996.

[20] I. Nikoladis, R. M. Fujimoto and A. Cooper A., “Time Parallel Simulation of Cascaded Statistical Multiplexes”, *Proceedings of 1994 ACM SIGMETRICS Conference on measurement and Modelling of Computer Systems*, May 1994.

[21] I. Nikoladis, R. Fujimoto and A. Cooper, “Parallel Simulation of High-Speed Network Multiplexers”, *Proceedings of the Conference on Decisions and Control* , Texas, December 1993.

[22] P. Heidelberger, “Fast Simulation of Rare Events in Queuing and Reliability Models”, *ACM Transaction on Modelling and Computer Simulation*, Vol. 5, No1, January 1995.

[23] Shahabuddin P.,”Fast Simulation of Packet Loss Rate in Communication Networks with Priorities”, *Proceedings of the Winter Simulation Conference*, 1994.

[24] M. Villen-Altamirano and J. Villen-Altamirano, “RESTART: A Straight Forward Method for Fast Simulation of Rare Events”, *Winter Simulation Conference Proceedings*, 1994.

[25] J. Schormans, E. Liu, L. Cuthbert and J. Pitts, “A Hybrid Technique for the Accelerated Simulation of ATM Networks and Network Elements”, *ACM Transaction on Modelling and Computer (TOMACS*) Vol. 11, Issue 2, April 2001.

[26] M. MacDougall, “Computer System Simulation: An Introduction*”, Computing Survey* 2, 3, Sept 1970.

[27] E. Lie, J. A. Schormans, L. Cuthbert and G. Stoneley, “A novel fast simulation method for ATM Network”, *Proceeding of IEEE Conference on High Performance Switching and Routing* 2000.

[28] J. A. Schormans, J.M Pitts, E.M. Scharf, A.J. Pearmain and C.I. Phillips, “Buffer Overflow Probability for Multiplexed on-off VoIP Sources”, *IEE Electronic Letters*, March 2000, Vol. 36. No. 6.

[29] J. Schormans, J. Pitts and L. Cuthbert, “Exact Fluid-Flow Analysis of Single on/off Source Feeding an ATM buffer”, *IEE Electronics Letters,* 1994, 30, (14), pp.1116-1117

[30] J. Schormans, E. Liu, R. Stewart and L. Cuthbert, “Analytical Technique for Accelerating the Simulation of Packet Networks,” *IEE Proc. Comm*., Volume 150, (5), 2003, pp. 341-346.

[31] A. H. I. Ma, *Accelerated Simulation of Power-Law traffic in Packets Networks*, PhD Thesis, Queen Mary University of London, 2003

[32] J. M. Pitts and J. A. Schormans, *Introductions to IP and ATM Design and Performance'*, John Wiley and Sons, Chichester, 2000.

[33] T. Bonald, A. Proutiere and J. W. Roberts, “Statistical Performance Guarantees for Streaming Flows Using Expedited Forwarding”, *IEEE INFOCOM*, 2001, 2, pp. 1104- 1112.

[34] O. Gurewitz, M. Sidi and I. Cidon, “The Ballot Theorem Strikes Again: Packet Loss Process Distribution”, *IEEE Transaction, on Information Theory*, 2000, 46, (7),

[35] J. A. Schormans, *Discrete Time Priority Queues in Telecommunication Networks*, PhD Thesis, Queen Mary University of London, 1990.

[36] J. A. Schormans and J. M. Pitts, “Analysis of Telecommunications Switch Model (GEO/D/1) with Time Priorities”*, Electronics Letters*, 1990, 26, (5), pp.325-326.